AFHRL-TR-87-69

AIR FORCE

HUMAN RESOURCES

AD-A196 349

DTIC FILE COPY

TERRAIN ACCURACY EFFECTS ON
SIMULATED RADAR IMAGE QUALITY

Peter M. Crane
Kevin W. Dixon, 1st Lt, USAF

OPERATIONS TRAINING DIVISION
Williams Air Force Base, Arizona 85240-6457

May 1988

Final Report for Period October 1986 - October 1987

LABORATORY

AIR FORCE SYSTEMS COMMAND
BROOKS AIR FORCE BASE, TEXAS 78235-5601

NOTICE

HAROLD G. JENSEN, Colonel, USAF
Commander

Unclassified

SECURITY CLASSIFICATION OF THIS PAGE

REPORT DOCUMENTATION PAGE

Form Approved OMB No 0704-0183

1a REPORT SECURITY CLASSIFICATION Unclassified	1b RESTRICTIVE MARKINGS
2a SECURITY CLASSIFICATION AUTHORITY	3 DISTRIBUTION/AVAILABILITY OF REPORT Approved for public release; distribution is unlimited.
2b DECLASSIFICATION/DOWNGRADING SCHEDULE	
4 PERFORMING ORGANIZATION REPORT NUMBER(S) AFHRL-TR-87-69	5 MONITORING ORGANIZATION REPORT NUMBER(S)
6a NAME OF PERFORMING ORGANIZATION Operations Training Division	6b OFFICE SYMBOL (If applicable) AFHRL/OT
7a NAME OF MONITORING ORGANIZATION	
6c. ADDRESS (City, State, and ZIP Code) Air Force Human Resources Laboratory Williams Air Force Base, Arizona 85240-6457	7b ADDRESS (City, State, and ZIP Code)
8a NAME OF FUNDING/SPONSORING ORGANIZATION Air Force Human Resources Laboratory	8b OFFICE SYMBOL (If applicable) HQ AFHRL
9 PROCUREMENT INSTRUMENT IDENTIFICATION NUMBER	
8c ADDRESS (City, State, and ZIP Code) Brooks Air Force Base, Texas 78235-5601	10 SOURCE OF FUNDING NUMBERS

PROGRAM ELEMENT NO	PROJECT NO	TASK NO	WORK UNIT NO
62205F	1123	33	01
	6114	33	01

11 TITLE (Include Security Classification)
Terrain Accuracy Effects on Simulated Radar Image Quality

12 PERSONAL AUTHOR(S)
Crane, P.M., Dixon, K.W.

13a TYPE OF REPORT	13b TIME COVERED	14 DATE OF REPORT (Year, Month, Day)	15 PAGE COUNT
Final	FROM Oct 86 TO Oct 87	May 1988	18

16 SUPPLEMENTARY NOTATION

17 COSATI CODES

FIELD	GROUP	SUB-GROUP
05	08	
05	09	

18 SUBJECT TERMS (Continue on reverse if necessary and identify by block number)
digital radar landmass simulator (DRLMS), radar images, transformation
image quality, simulation, accuracy,
radar mapping, terrain accuracy

19 ABSTRACT (Continue on reverse if necessary and identify by block number)

The effect of varying the degree of accuracy in the transformation of terrain elevation data into simulated real beam ground mapping radar was assessed. In two experiments, Air Force navigators judged the quality and training value of simulated radar imagery of a mountainous area produced at six levels of transformation accuracy. The results show that (a) increasing transformation accuracy above the simulator's current standard did not significantly increase either the judged quality of the imagery or its perceived training value; (b) decreasing terrain vertical accuracy below the current standard significantly decreased both judged quality and perceived training value.

20 DISTRIBUTION/AVAILABILITY OF ABSTRACT [X] UNCLASSIFIED/UNLIMITED [ ] SAME AS RPT [ ] DTIC USERS	21 ABSTRACT SECURITY CLASSIFICATION
22a NAME OF RESPONSIBLE INDIVIDUAL Nancy J. Allin, Chief, STINFO Office	22b TELEPHONE (Include Area Code) (512) 536-3877
22c OFFICE SYMBOL AFHRL/TSR	

DD Form 1473, JUN 86 Previous editions are obsolete

SECURITY CLASSIFICATION OF THIS PAGE

Unclassified

SUMMARY

This effort was conducted to determine if more accurate transformations of digital terrain elevation data would significantly increase the perceived quality or training effectiveness ratings of simulated real beam ground mapping radar imagery. Although an increase in accuracy produces a more detailed image, it also requires more computer time to generate and therefore increases the cost of data base development.

Seven KC-135 and 25 C-130 navigators evaluated simulated radar images generated at six different levels of transformation accuracy for usefulness in navigation and in training navigators. Analyses show that transformations which were more accurate than current standards did not produce perceptible increases in image quality or in rated training effectiveness. Transformations which were less accurate than current standards were readily discriminated and judged to be of poorer quality. It is recommended that the current standard for vertical transformation accuracy be maintained for future radar landmass simulators.

PREFACE

This project was conducted in support of the Air Force Human Resources Laboratory's Technical Planning Objective, Aircrew Training Technology. The goal of this effort is to develop cost-effective strategies and equipment for aircrew training. This experiment was conducted under Work Units 6114-33-01, Analysis of Imagery for Manipulation, and 1123-33-01, Fidelity Requirements for Sensor Imagery.

The authors wish to thank John Stengel and Capt John Dooley of the Aeronautical Systems Division/Engineering for providing the stimuli; and the navigators of the 161st Air Refueling Group, Arizona Air National Guard, and the 34th Tactical Training Group, Little Rock AFB, Arkansas, who participated as subjects.

The authors also thank Ms. Eileen M. Evans for all the time and effort she spent in helping to prepare this report.

TABLE OF CONTENTS

LIST OF FIGURES

TERRAIN ACCURACY EFFECTS ON SIMULATED RADAR IMAGE QUALITY

I. INTRODUCTION

The accuracy of a simulated ground mapping radar image in polygon-based radar simulators is determined, in part, by the number of polygons or faces used to model the earth's terrain. Experiments were conducted to determine if a more detailed transformation of digital terrain elevation data would significantly increase the training effectiveness of simulated radar imagery. The results failed to show any significant increase in the perceived image quality or the perceived training value of the simulated radar imagery produced from the more detailed transformations.

Many current digital radar landmass simulators (DRLMSs) rely on digital data provided by the Defense Mapping Agency for terrain elevation information. These data typically consist of terrain elevations above mean sea level (MSL) at intervals of 100 meters. A DRLMS transformation program fits polygons around these elevation values to depict the terrain. The transformed terrain data are then merged with surface feature information and stored for use in the real-time simulation.

One variable in the elevation data that influences the transformation program is terrain roughness. For perfectly flat ground, roughness equals zero whereas for mountains, roughness can be greater than 10. In effect, roughness measures the rate of elevation change between data posts in a geographic region. The transformation program fits polygons about the elevation posts using an iterative process in which the simulated terrain is tested against a criterion measuring transformation accuracy. This accuracy criterion is inversely related to roughness. As roughness increases, terrain accuracy decreases; allowable differences between elevations in the source data and elevations of the transformed data increase. Simulated terrain elevations are, therefore, highly accurate for flat terrain and less accurate for rough or mountainous areas. This accuracy criterion influences the number of polygons and, therefore, the fidelity or realism of the simulated radar scene. As the elevation data are transformed, additional polygons are added in an iterative manner until the specified criterion is met. Because of the iterative transformation process the accuracy criterion directly influences the number of polygons and the length of time required to transform the digital terrain data. Since the accuracy criterion influences the length of time required to complete the transformation process, it also affects the cost of developing the simulation. Therefore, radar simulations of the same area produced using differing accuracy criteria can differ significantly in both fidelity and cost.

Because fidelity and cost are both directly related to the accuracy criterion, it is important to determine the relation between the fidelity of the simulated radar scene and its training effectiveness. As Roscoe (1980) pointed out, beyond a certain point, increases in simulation fidelity produce little or no increase in training. The purpose of the present experiments was to determine the relation between six levels of transformation accuracy and (a) the perceived similarity of the six simulated radar images, and (b) the perceived training value of those images.

A modified KC-135 DRLMS was used to transform and display simulated ground mapping radar imagery of an area 40 miles southeast of Knoxville, Tennessee. Six different accuracy criteria were used. Photographs of these images were evaluated by Air Force navigators. Images were evaluated on their similarity, their overall image quality, and their perceived value for training. The six levels of transformation accuracy used included the current standard for this DRLMS, two more accurate transformations, and three less accurate transformations. In Experiment 1, seven KC-135 navigators evaluated simulations produced for 10,000' MSL (approximately 6,000' above ground level [AGL]) at 15 and 30nm. In Experiment 2, 25 C-130 navigators evaluated images of the same area produced for 5,000' MSL (approximately 1,000' AGL) and at 15 and 20nm.

II. EXPERIMENT 1

Research Methods

Apparatus

A KC-135 DRLMS modified for engineering development was used to transform and display simulated radar images of an area 40 miles southeast of Knoxville, Tennessee. Six separate transformations of the digital terrain elevation data for this area were produced using different accuracy criteria. With the apparatus available, it was not possible to directly manipulate transformation accuracy. It was possible, however, to change transformation accuracy indirectly by altering the specified terrain roughness criterion. Since the transformation accuracy requirement decreases with increasing roughness, multiplying terrain roughness by weights less than one has the effect of increasing transformation accuracy. Likewise, multiplying roughness by weights greater than one produces less accurate transformations. The weights used in this experiment were 6, 4, 2, 1, 0.8 and 0.6. The terrain models generated by transforming the elevation data were used to produce simulated radar images at 15 and 30nm ranges. Only terrain data were used to generate these simulations. Surface features such as vegetation, water, or structures were not included. The images were equivalent to ground mapping radar from a KC-135 flying at 10,000 feet above mean sea level (approximately 6,000' AGL). Glossy, black-and-white photographs were obtained for each of the 12 simulated radar images. Figures 1 through 3 are examples of different transformations of the 30nm imagery. Figure 1 was produced with a weight of 0.6; this image is more accurate than Figure 2, which is the current standard with a weight of 1. Figure 3 shows the least accurate transformation, with a weight of 6.

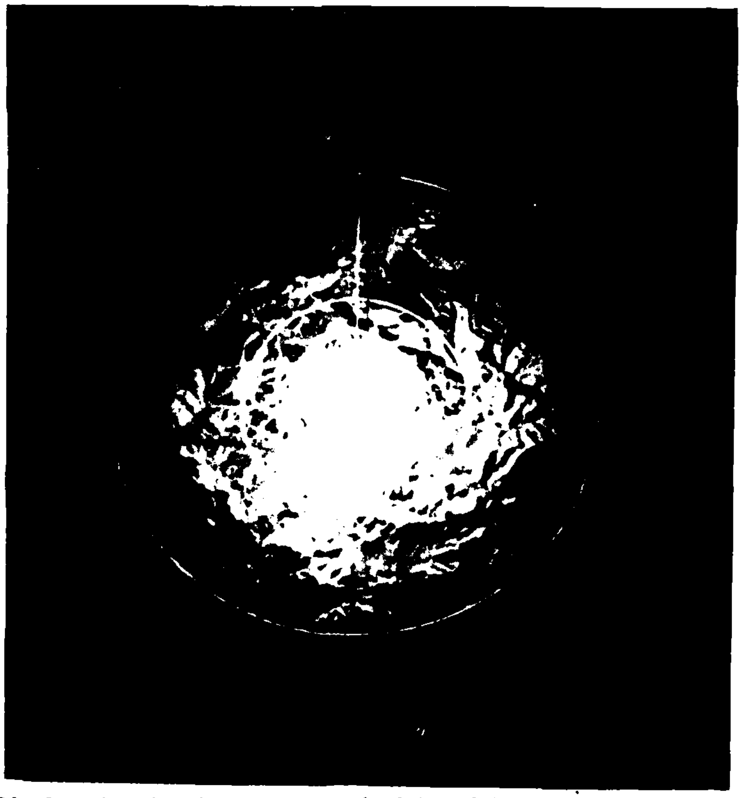

Figure 1. **Simulated radar image produced with weight = 0.6 for Experiment 1; this is the most accurate transformation in the stimulus set. Altitude is 10,000 feet MSL; range is 30nm; north is at the top. This image was produced from terrain elevation data only and does not contain any surface features.**

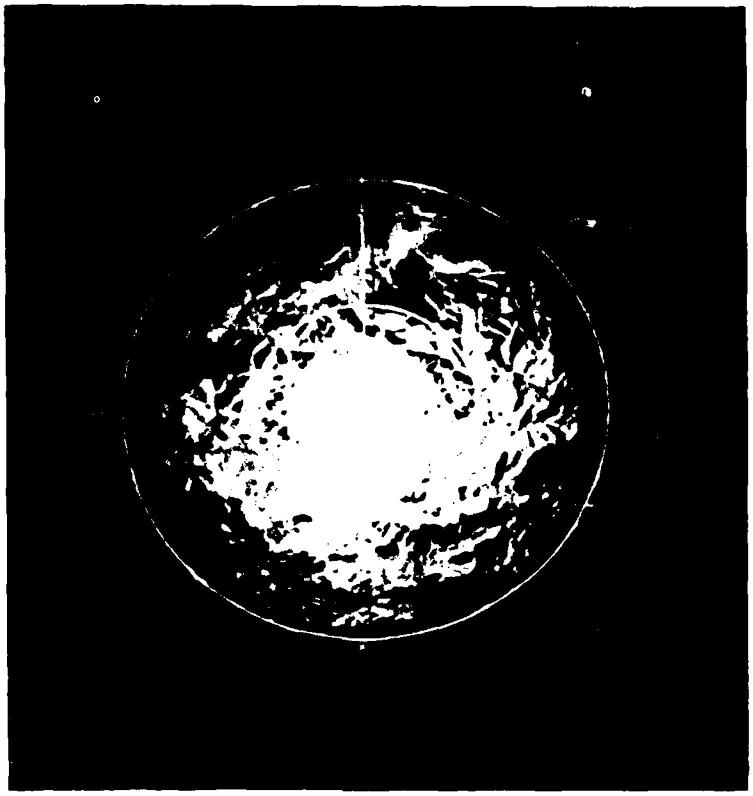

Figure 2. Simulated radar image produced with weight = 1, standard accuracy, for Experiment I.

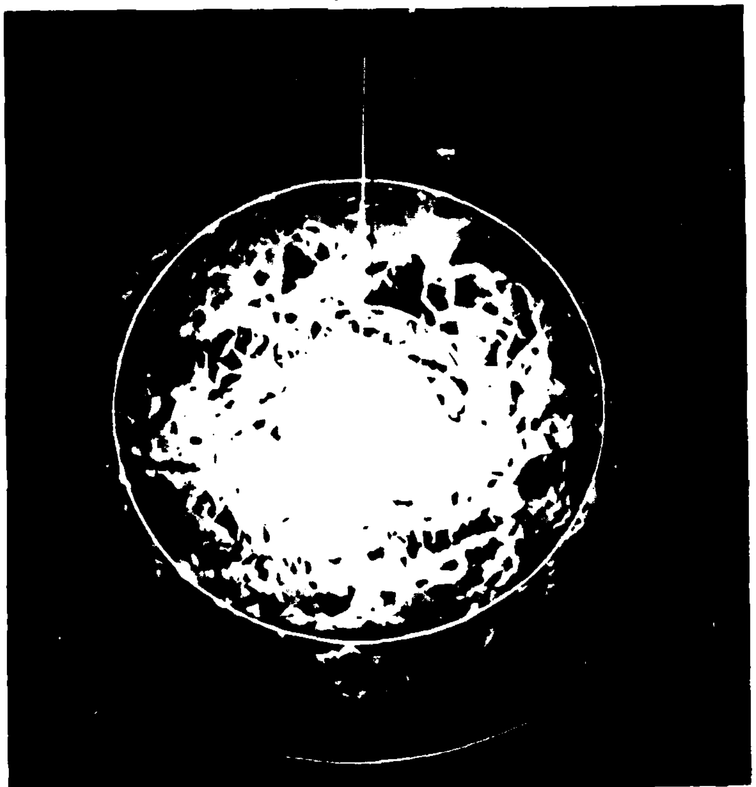

Figure 3. Simulated radar image produced with weight = 6 for Experiment 1. This is the least accurate transformation in the stimulus set.

Subjects

Seven KC-135 navigators from the 161st Air Refueling Group, Arizona National Guard, participated in Experiment 1. Each subject was operationally qualified and current.

Procedure

Subjects were tested individually and asked to judge the quality of the simulated radar imagery. Each navigator was briefed on the purpose of the experiment and given a 1:50,000 navigation chart of the area, a pair of dividers, and a 30nm simulated radar image with an accuracy criterion of 1.0 (Figure 2). The navigators were then asked to orient themselves and to identify several prominent terrain features to ensure they were able to correctly interpret the image. All subjects were able to do this without difficulty.

Following this introduction procedure, each navigator evaluated the quality of the 30nm simulated imagery by performing three tasks: preference, ranking, and rating. The three tasks were then repeated for the 15nm range. All tasks were self-paced and took approximately 30 minutes to complete. After data collection, subjects were asked if they could offer any comments or suggestions about the experimental procedures or materials.

In the preference task, each image for a specific range was paired with each of the other images for that range. Each pairing was presented twice in random order for a total of 30 judgments; the order of presentation of the images was counterbalanced. The subjects were instructed to indicate which image in each pair they would rather use in navigation tasks and to indicate a preference even if differences between the images were very slight. After completing the preference task, subjects ranked the six images from best to poorest quality. To perform this task, subjects were given all six images and instructed to arrange them until they were satisfied with the order.

It was possible that even images judged to be poorer in quality than others might be considered adequate for training. Likewise, it was possible that simulated images ranked as better than all others might be judged inadequate for effective training. To test this, subjects were also asked to rate each image on a 5-point scale of "Usefulness for Training Navigators." The scale anchors were "1. Excellent, indistinguishable from aircraft radar," and "5. Unacceptable, not useful for training."

Results

The preference data were analyzed using a Thurstone scaling approach that converted the data to an equal-interval scale of perceived quality (Baird & Noma, 1978, Chapter 7; Nunnaly, 1978, Chapter 2). The zero point of this scale is arbitrary, and the interval between images is based on standard deviations computed from the unit normal distribution. Figure 4 shows that for the 30nm range, the three most accurate transformations (weight = 1, 0.8, and 0.6) form a tight cluster. This clustering indicates that these three images are not discriminably different from each other. The results for the 15nm range are different. They show that the two most accurate transformations, weights of 0.8 and 0.6, are not discriminable from each other, and that the third most accurate weight of 1 is discriminably poorer than the two more accurate transformations.

The mean ranking data are summarized in Figure 5. A Friedman Multi-Sample Test (Bradley, 1968, Chapter 5) indicates no significant differences among the mean ranks for the three more accurate transformations (alpha = .06) at either range. However, the differences between the mean ranks for these transformations and the next transformation level, 2, were significant ($\underline{p} < .05$) for both ranges. These results indicate that the three most accurate transformations were not perceptibly different from each other and that the next level was consistently ranked as lower in quality.

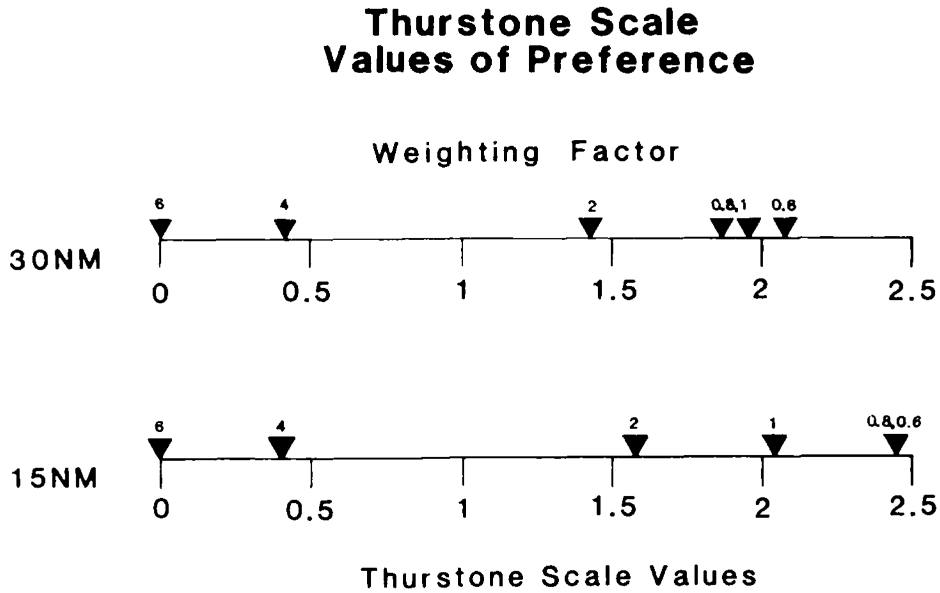

Figure 4. Thurstone scales of preference data for Experiment 1. Higher scale values indicate greater preference. Lower transformation accuracy weights indicate greater fidelity.

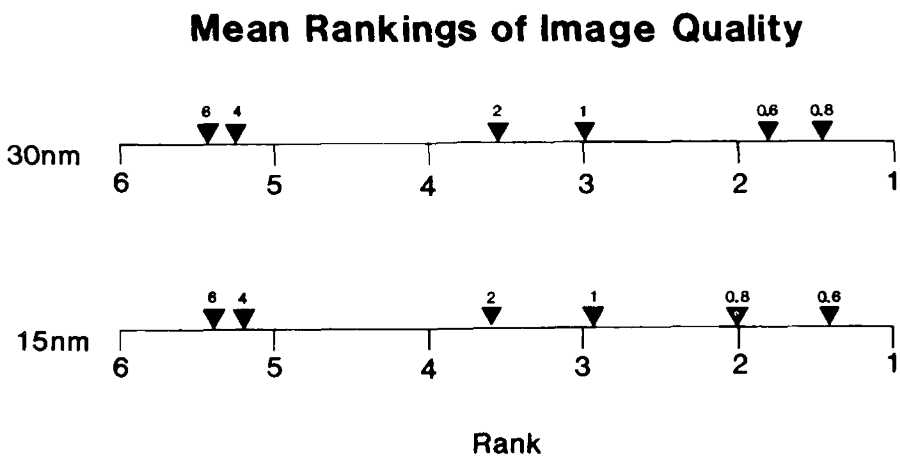

Figure 5. Mean rankings of image quality for Experiment 1.

The mean rated training value of each image is shown in Figure 6. These data were analyzed using a one-way analysis of variance (Winer, 1971, Chapter 4). These analyses show that the ratings assigned to the three most accurate transformations were not significantly different from each other for either range (alpha = .05) and that the mean rating for these images was significantly higher (p .05) than the next most accurate transformation level.

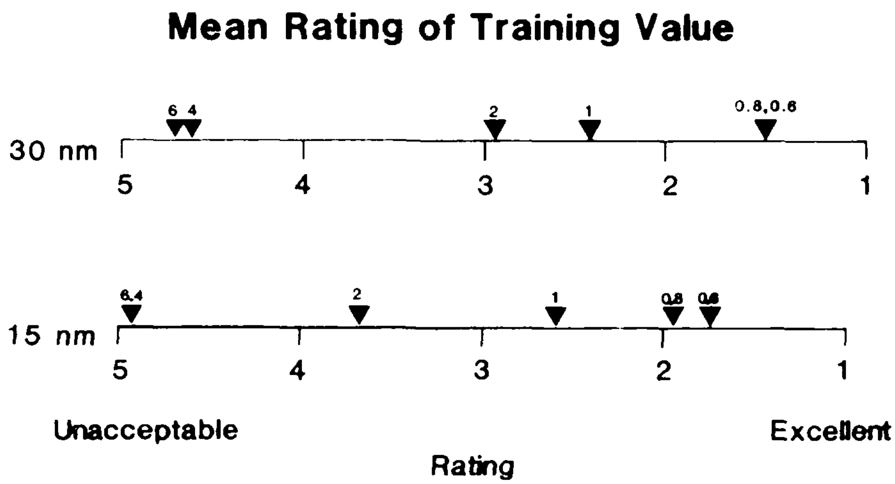

Figure 6. Mean ratings of training value for Experiment 1.

Discussion

Experiment 1 addressed two major questions. The first question dealt with the level of fidelity or realism that is necessary to produce an adequate simulation. In this study, increasing fidelity was achieved by increasing terrain vertical accuracy; i.e., changing the accuracy criterion. Increasing transformation accuracy significantly increases the processing time necessary for the DRLMS to transform digital elevation data into a simulated radar image. Compared with the standard transformation (weight = 1), the most accurate transformation (weight = 0.6) required 25% more processing time while the least accurate (weight = 6) required 39% less time. Since transformation of terrain data requires several hours of processing per 2° x 2° cell, the cost differences are significant. The data obtained from the preference and ranking tasks indicate that transformations with weights of 1, 0.8, and 0.6 are perceptually similar to one another. Therefore, it can be concluded that increasing the fidelity of the simulation by using a weighting factor less than 1.0 produces no significant gain in the perceived quality of the simulated image.

The second question addressed by Experiment 1 concerns training effectiveness. The preference and ranking tasks assess the discriminability of the stimuli but not their usefulness for training. The rating data, however, show that the three more accurate transformations are judged to have equal training value and to be significantly more useful than the less accurate images. According to Roscoe (1980), increasing realism beyond some point can increase cost without increasing training effectiveness. These results support Roscoe's assertion. For both ranges, radar simulations produced with normal accuracy, weight = 1, (a) are indistinguishable from more accurate simulations with weights of 0.8 and 0.6, (b) have the same training value as the more accurate transformations, and (c) can be produced at significantly lower cost. Although the training effectiveness of each level of simulation fidelity cannot be tested directly, a reasonable assumption is that if operational navigators cannot reliably distinguish between the various levels of transformation accuracy, then these simulations should have approximately the same training value.

Although the results of Experiment 1 seem clear, the navigators who served as subjects consistently mentioned the altitude of the simulations, 10,000' MSL (approximately 6,000' AGL), as a problem. These subjects were navigators on KC-135 aircraft, which normally fly at altitudes

above 30,000 feet. Other aircraft, such as C-130, B-52, or FB-111, often fly at less than 1,000 feet above the highest local terrain. The stimulus images produced at 10,000 feet are therefore not representative of most Air Force navigation environments. Since the effect of transformation accuracy on image quality is more pronounced at low altitude than at high altitude, a level of transformation accuracy which produces acceptable images for low altitudes will also produce acceptable images for higher altitudes. Experiment 2 was conducted to validate these results using low-altitude simulations.

III. EXPERIMENT 2

Research Methods

Apparatus

Simulated ground mapping radar images were produced using the same DRLMS and data base transformations as in Experiment 1. Altitude, however, was reduced to 5,000' MSL (approximately 1,000' AGL) and ranges changed to 15 and 20nm. Figures 7 through 9 are examples of the images used in Experiment 2.

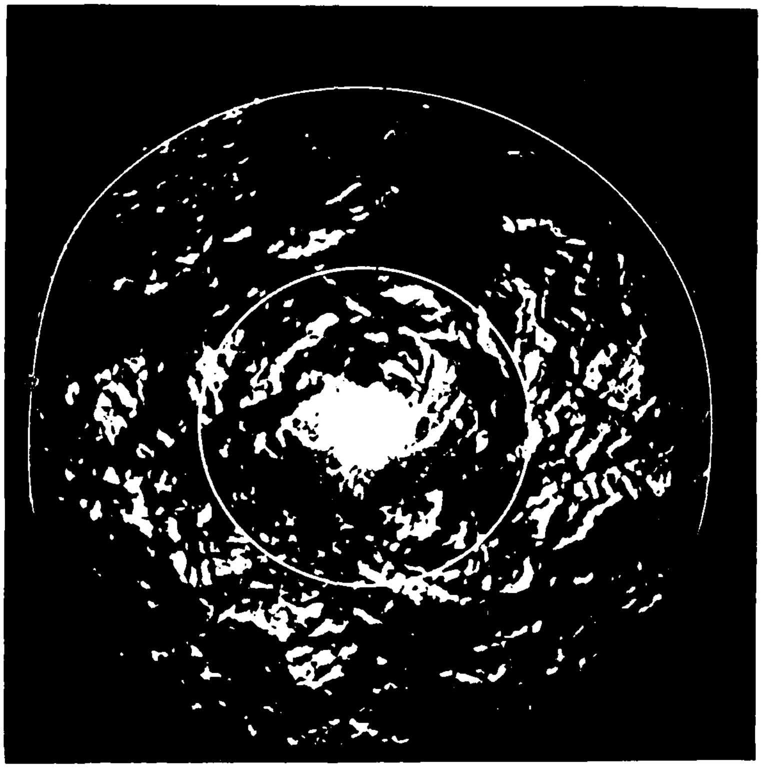

Figure 7. **Simulated radar image produced using standard accuracy for Experiment 2; weight = 1. Altitude is 5,000' MSL; range is 20nm; north is at the top.**

Subjects

Twenty-five C-130 navigators from the 34th Tactical Training Group, Little Rock AFB, Arkansas, participated in Experiment 2.

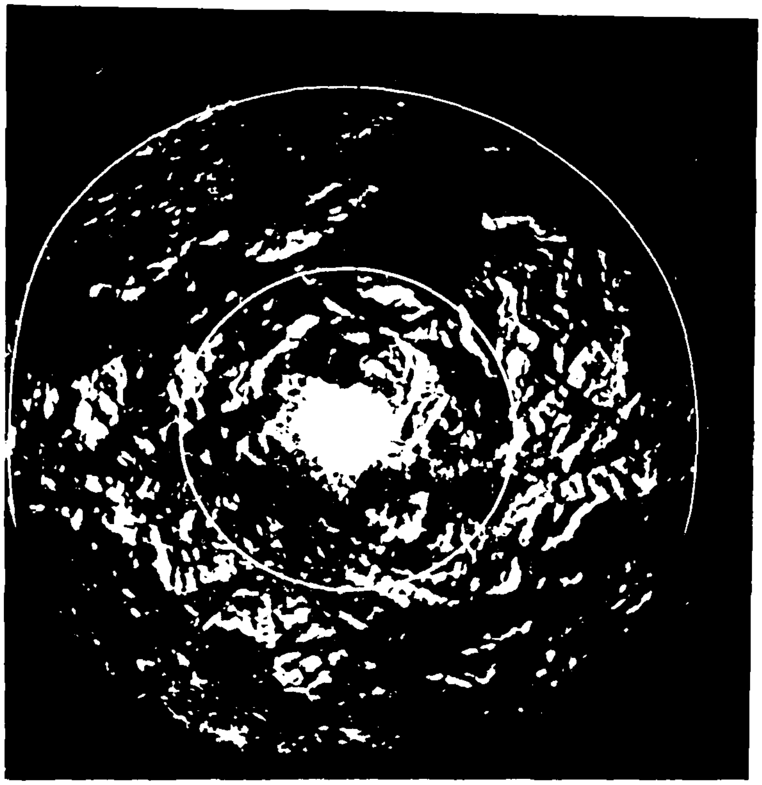

Figure 8. Simulated radar image with weight = 0.6 for Experiment 2. This is the most accurate transformation.

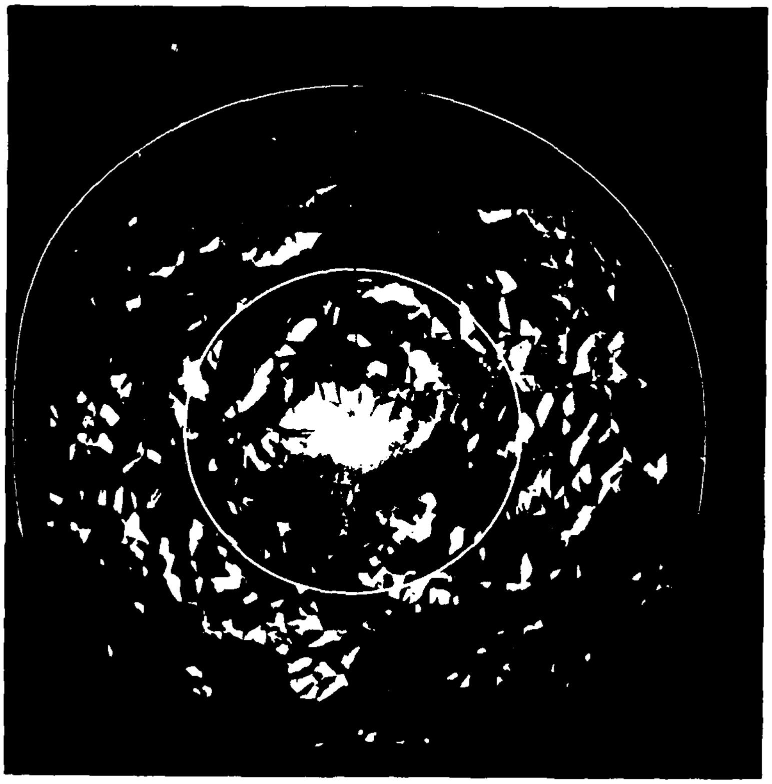

Figure 9. Simulated radar image with weight = 6 for Experiment 2. This is the least accurate transformation.

Procedure

Subjects completed the three tasks--preference, ranking, and rating--as in Experiment 1.

Results

Thurstone scale analysis of the preference data is shown in Figure 10. The analysis shows that for the 20nm range, the three most accurate transformations (weights equal to 0.6, 0.8, and 1) form a cluster. This clustering indicates that these transformations are not highly distinguishable from each other. The interval between 1 and the less accurate transformations is larger than the interval between 1 and the more accurate transformations, indicating larger perceptual differences. Thurstone scaling analysis for the 15nm range shows an even tighter clustering for the three most accurate transformations; however, the order is reversed from that for the 20nm range. A forced choice procedure was used with the preference task so that evaluators were required to specify a choice even if perceived differences between images were minimal. If two images are truly indistinguishable, evaluators will respond at random and rate each image as preferred 50% of the time; in this case, the interval between images on a Thurstone scale would be zero. This appears to be the case for the three most accurate transformations at the 15nm range; the images were perceptually indistinguishable from each other, and the apparent reversal is the result of random variation in responses. The interval between this cluster and the scale values for the less accurate transformations is much larger than the intervals within the cluster, indicating larger perceived differences.

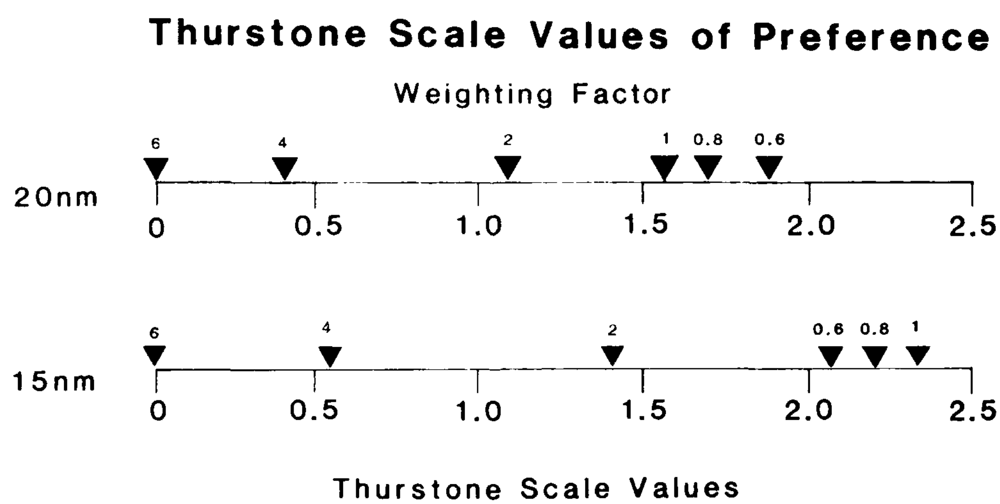

Figure 10. **Thurstone scales of preference data for Experiment 2.**

The mean ranking data are summarized in Figure 11. A Friedman Multi-Sample Test indicates no significant difference between the mean ranks assigned to the two highest ranked transformations, 0.6 and 1, for the 20nm range (alpha = .05). However, the difference between the ranks assigned to these images and the rank assigned to the next image, 0.8, was significant ($p < .05$). For the 15nm range, no significant differences were found between the ranks assigned to the three most accurate transformations: 1, 0.8, and 0.6. There was a significant difference between the ranks assigned to these transformations and the next transformation level, 2 ($p < .05$). These results indicate that the two highest ranked transformations for the 20nm range and the three highest

ranked transformations for the 15nm range were not perceptibly different from each other and that the next ranked image was consistently perceived as lower in overall quality.

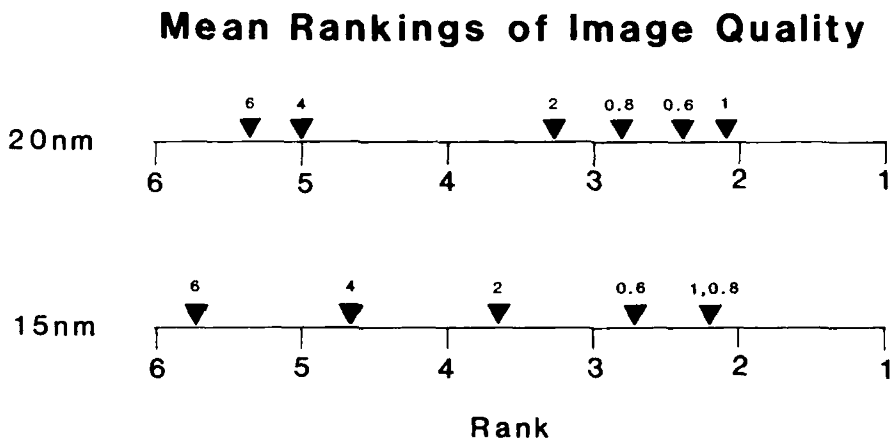

Figure 11. **Mean rankings of image quality for Experiment 2.**

The mean rated training values are shown in Figure 12. These data were analyzed using a one-way analysis of variance. These analyses show that the ratings assigned to the two highest rated transformations (1 and 0.6) were not significantly different from each other for the 20nm range and that the ratings assigned to the three highest rated transformations (1, 0.8, and 0.6) were not significantly different from each other for the 15nm range (alpha = 0.5). The mean ratings for these transformations were significantly higher than the rating for the next lower transformation level ($p < .05$).

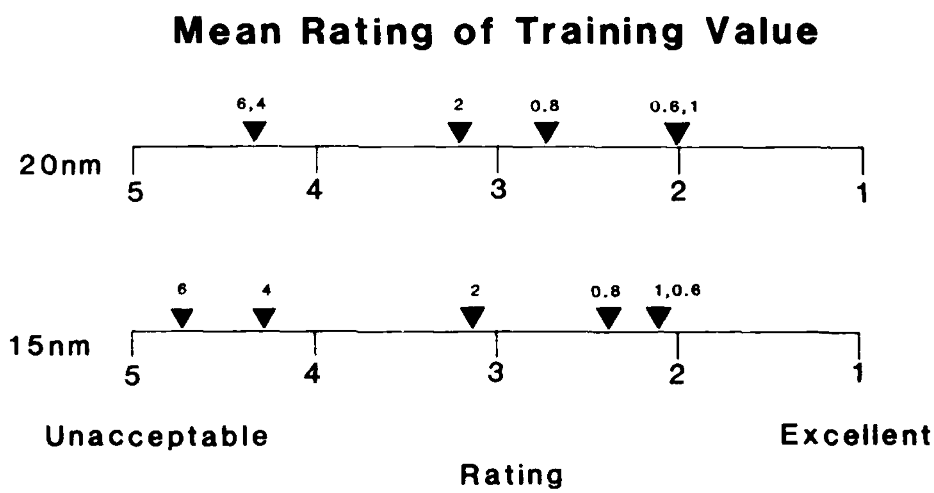

Figure 12. **Mean ratings of image quality for Experiment 2.**

Discussion

Overall, the data from Experiment 2 support the conclusions from Experiment 1. The data from the preference and ranking tasks for the 15nm range indicate that transformations with weights of 1, 0.8, and 0.6 are perceived as highly similar to each other. For this range, it can be concluded that increasing the fidelity of the simulation by increasing the accuracy criterion beyond the current system standard will produce no significant gains in the perceived quality of the simulated images.

The data for the 20nm range are not as consistent. The Thurstone scale values (i.e., preference data) for the three most accurate images form a cluster, although not as closely as for the 15nm range. Furthermore, the intervals between the three most accurate transformations are less than the intervals between the current standard (weight = 1) and the less accurate image (weight = 2). These data show that the perceived differences between the current standard and the more accurate images are small compared to differences between the current standard and less accurate images. The ranking data show the same pattern of results for all weights except for 0.8, which is ranked significantly lower than the current standard. The rating data show that for the 15nm range, the three more accurate transformations are judged to have equal training value and to be significantly more useful than the less accurate images. This pattern of results is also true for the 20nm range except for 0.8, which is rated significantly lower. For the 20nm, 0.8 simulated radar image that was used for both ranking and rating tasks, no obvious flaw can be found which would produce lower rankings or training effectiveness ratings. However, since these results are in such disagreement with the results for the 15nm range and from Experiment 1, the data must be considered suspect and will not be included in further discussions.

IV. CONCLUSIONS

The data collected in these two experiments support the following conclusions:

1. Simulated images produced from transformations which are less accurate than the current DRLMS system standard can be readily distinguished from standard imagery and are rated as providing significantly lower training value.

2. Simulated images produced from transformations which are more accurate than the current system standard cannot be reliably distinguished from standard imagery and are rated as having the same training value.

3. The procedures used in this research have identified a point at which increases in simulation fidelity do not produce increases in perceivable image quality or in rated training value; the only result is increased cost.

It is recommended that the current standard for vertical terrain transformation accuracy or its equivalent be retained in specifications for future radar landmass simulators.

REFERENCES

Baird, J.C., & Noma, E. (1978). Fundamentals of scaling and psychophysics. New York: John Wiley & Sons.

Bradley, J.V. (1968). Distribution-free statistical tests. Englewood Cliffs, NJ: Prentice-Hall.

Nunnaly, J. (1978). Psychometric theory (2nd ed.). New York: McGraw-Hill.

Roscoe, S.N. (1980). Transfer and cost effectiveness of ground-based flight trainers. In S.N. Roscoe (Ed.), Aviation Psychology. Ames, IA: Iowa State University Press.

Winer, B.J. (1971). Experimental principles in experimental design (2nd ed.). New York: McGraw-Hill.

U.S. GOVERNMENT PRINTING OFFICE: 1988 561-048/80020